

Autor / Author

GARRIDO MERCHÁN, Eduardo César

Universidad Pontificia Comillas

<https://doi.org/10.32466/eufv-rel.2022.9.744.109-126>

RECIBIDO / RECEIVED 3 de octubre de 2022

ACEPTADO / ACCEPTED 25 de octubre de 2022

PÁGINAS / PAGES De la 109 a la 126

ISSN / ISSN 2386-2912

Transhumanismo y consciencia fenoménica

Transhumanism and phenomenal consciousness

Los últimos avances de la ciencia invitan a pensar en la posibilidad práctica del transhumanismo. Pero quizá es pertinente estudiar una pregunta a la que todavía ni la ciencia ni la filosofía han sabido dar respuesta. En concreto, la naturaleza de la consciencia fenoménica de Ned Block o del concepto de vigilancia de Stanislas Dehaene. Pese a los últimos avances en neurociencia y en inteligencia artificial, seguimos sin pistas para responder a la pregunta del problema fuerte de la consciencia de David Chalmers, de cómo y por qué estamos aquí. Somos capaces de discriminar de qué información somos conscientes, pero no hemos conseguido averiguar por qué somos conscientes o cómo se genera esa consciencia. El problema es más grave de lo que parece, ya que no somos capaces de medir, definir ni replicar la consciencia fenoménica. Sin medición, si el objeto de estudio es subjetivo y cualitativo, no se puede aplicar el método científico. Luego, pese a que podríamos replicar la información de nuestro cerebro en un artefacto mecánico, nada garantiza que nuestra consciencia fenoménica se replique ahí, ya que no podemos medir la consciencia fenoménica. Este artefacto hecho a imitación nuestra podría no ser más que un *zombie* filosófico, en palabras de David Chalmers, un ser sin consciencia fenoménica, un ser sin alma. Ante esta realidad, cabe preguntarnos: ¿es seguro explorar la alternativa de trascender en un sistema independiente de nuestro cerebro biológico, como un ordenador, si no tenemos ninguna herramienta para medir si ese sistema puede ser consciente de sí mismo o producir consciencia fenoménica como un epifenómeno? Si la respuesta, como se defiende en este trabajo, es negativa, una alternativa más segura para el transhumanismo es explorar la ingeniería genética, los implantes artificiales o embeber el cerebro en sistemas artificiales.

#consciencia fenoménica, #transferencia de consciencia a un ordenador, #transhumanismo, #misterianismo.

Science latest advances invite to think in the practical possibility of transhumanism. However, it may be a prerequisite to study a question that has not been answered yet by science nor philosophy, the hard problem of consciousness postulated by David Chalmers: why and how are we here. We are able to scientifically determine what information are we conscious, although we have not managed to know why or how are we conscious. The problem is really hard, as we are unable to measure, define nor replicate phenomenal consciousness. Without a measure, if the object of study is subjective and cualitative, we can not apply the scientific method. Therefore, although we could replicate the information of our brain in an artificial device, anything guarantees that our phenomenal consciousness would be replicated there, as we can not measure phenomenal consciousness. This artifact, that imitates the human being, may be a philosophical zombie, a being without consciousness, without soul. Thinking about this fact, we may wonder: is it safe to explore the alternative of trascending in a system that is independent from our biological brain, as a computer, if we do not have any tool to measure whether this system can be conscious of itself or to produce phenomal consciousness as an epiphenomenon? If the answer, as in this work is defended, is negative, a safer alternative for transhumanism may be to explore genetic engineering, artificial enhancements or embed the brain in artificial systems.

#phenomenal consciousness, #consciousness transfer to a computer, #transhumanism, #mysterianism.

1. El problema fuerte de la consciencia fenoménica

La consciencia, o el alma o el ser, ha sido objeto de estudio por el ser humano desde la Antigüedad (Copleston, 2003). En la psicología popular, se le ha referenciado muchas veces como el alma o el espíritu (Bering, 2006). Sin embargo, esta definición ha precisado de un mayor rigor científico para poder ser estudiada objetivamente. No obstante, al ser un fenómeno cualitativo, subjetivo y relativo al observador, uno de los primeros problemas que surgen al tratar de estudiar la consciencia es aportar una definición teórica desde el punto de vista epistemológico. En palabras de Juan Arana, para definir la consciencia debemos o bien ser claros o por otro lado ser precisos, pero, por desgracia, parece imposible satisfacer ambos criterios simultáneamente. En este sentido, y siendo concretos, podemos ser claros definiendo la consciencia como la dimensión autotransparente de la mente, o precisos al definirla como el correlato ontológico de la introspección mental. En ambos casos, se descubre que la definición de consciencia es un fenómeno complejo para ser trabajado por el método científico, debido a sus particularidades. En palabras de Karl Popper, la consciencia se asemeja más a la nube

que a un reloj. Del reloj podemos medir objetivamente sus propiedades fácilmente pero para estudiar a la nube necesitamos otro tipo de enfoque distinto.

En ciencias de la computación, se recurre al principio «divide y vencerás» cuando un objeto de estudio es complejo. Siguiendo esta metodología, el filósofo de la mente Ned Block separó nuestra concepción de consciencia en consciencia fenoménica, la experimentación de los fenómenos que nos rodean y cómo los experimentamos como un sujeto subyacente a la burbuja de información interna y externa que experimentamos en cada momento, y consciencia de acceso o a qué sensaciones decide nuestro subconsciente prestar atención (Block, 1995). Adicionalmente, resulta útil añadir que también poseemos una consciencia de nuestra identidad, de nuestro yo narrativo o consciencia del yo, o del ego (Nelson, 2003). Pese a que gracias al método científico se puede provocar una disolución temporal de la consciencia del yo o muerte del ego, por ejemplo usando psicodélicos (Nour *et al.*, 2020) como el DMT (Timmermann, 2018), y también se puede determinar qué información determina el subconsciente a la que puede acceder la burbuja consciente (Dehaene, 2014), que es la consciencia de acceso, el método científico ha dicho muy poco, si es que ha aportado algo, de la consciencia fenoménica o, como también se denomina en neurociencia, la vigilancia, que será nuestro objeto de estudio en este trabajo.

Esta dificultad para trabajar con el fenómeno de la consciencia fenoménica, de nuestra percepción privada de la actividad cerebral debida a estímulos internos al cuerpo y externos que provienen a través de los sentidos, se hace palpable al hacernos preguntas sobre los fenómenos que experimentamos: ¿por qué experimentamos el rojo de esta forma determinada?, ¿cómo se puede definir la rojez del color rojo?, ¿cómo le podemos explicar a una persona ciega cómo es el color rojo? Si el lector hace en este momento un breve ejercicio de reflexión, descubrirá que es inefable aportar una descripción del color rojo para una persona que nunca ha visto el color rojo debido a un problema de visión. En este particular apartado encontramos el argumento del conocimiento de la habitación de Mary (Jackson, 2004). Mary es una experta en visión, pero es ciega desde su nacimiento. Por mucho conocimiento objetivo que Mary tenga de la visión desde un punto de vista objetivo o epistémico, Mary nunca conocerá cómo percibe el ser humano el color rojo, o el carácter ontológico del rojo. En otras palabras, no posee una consciencia fenoménica del color rojo ya que no recibe esta información por medio de sus sentidos y, en consecuencia, no experimenta su Quale. Este breve ejercicio mental sirve de evidencia de que la consciencia fenoménica es un conocimiento ontológicamente subjetivo y cualitativo (Searle, 1997), inefable y posiblemente inabordable por el método científico (Chalmers, 2002), tal y como se desarrollará a lo largo de este trabajo. Otro ejercicio mental para probar el carácter subjetivo, cualitativo y relativo al observador de la consciencia fenoménica es el siguiente: si se supone que, a través de la ciencia, se logra un conocimiento perfecto de la anatomía de un murciélago, ¿puede el método científico, con todo el conocimiento adquirido, hacernos sentir cómo se siente un murciélago o reproducir la experiencia consciente de un murciélago (Nagel, 1974)? Pese a adquirir un conocimiento pleno de las facultades del murciélago, a partir de un conocimiento objetivo de este no se encuentra ningún «fuego en las ecua-

ciones» que haga emerger la consciencia fenoménica ya que, para empezar, no hay ninguna forma de medir este concepto que no sea mediante correlaciones, de las cuales no se puede extraer ninguna noción de causalidad o conocimiento objetivo sobre el cómo y mucho menos el porqué de que la consciencia fenoménica emerja de, al fin y al cabo, la materia. Llegados a este punto, las únicas hipótesis plausibles parecen ser afirmar que la consciencia es un fenómeno fundamental e irreducible del universo, o bien, como algunos filósofos plantean, que no existe y es una ilusión. Sin embargo, esta segunda hipótesis parece descartarse ya que, al menos, el autor de este texto tiene máxima certeza con respecto a toda la realidad que le rodea y de lo único que está seguro es de que es consciente. Podemos hacernos más preguntas sobre la consciencia fenoménica de las que, por sus características, nunca obtendremos una solución objetiva: ¿puede explicar la ciencia cómo y por qué sentimos experiencias relacionadas con fenómenos internos y externos y por qué y cómo estas experiencias se traducen en determinadas características como colores y sabores? (Chalmers, 2002). Incluso en condiciones de aislamiento sensorial y de meditación, un sujeto es consciente de su existencia (Solomon *et al.*, 1957). El problema de cómo y por qué emerge la consciencia fenoménica se conoce como el problema difícil de la consciencia (Chalmers, 2007) y, si se razona sobre ello, al preguntarse el porqué de estas experiencias podemos equipararlo incluso con la pregunta metafísica del por qué existe algo y no nada (Krauss, 2013). Ya que si eliminamos nuestra experiencia consciente de la realidad que nos rodea, el universo no sería consciente de sí mismo, y eso puede ser equiparado, al menos desde nuestra experiencia individual, a que no exista nada. Sin embargo, al ser conscientes, somos conscientes de que existe algo (Descartes, 2013). La pregunta que resta resolver es cómo y por qué.

Este trabajo realiza un estudio de la consciencia fenoménica para aportar un argumento que pretende advertir sobre los riesgos a los que un individuo se puede exponer en el hipotético caso de que una corriente transhumanista pueda hacer que se desarrolle un método por el cual se afirme que el individuo puede transferir su consciencia fenoménica a una máquina. En primer lugar, estudiaremos la consciencia fenoménica desde el punto de vista del funcionalismo, corriente filosófica materialista que afirma que la consciencia surge de estados funcionales que representan los estados mentales de un sujeto. En esa sección se expondrá un breve estado del arte sobre modelos de computación e inteligencia artificial, desde el punto de vista de la ciencia de la computación, en los que se afirma que se puede hacer emerger consciencia en máquinas siguiendo postulados funcionalistas. Tras presentar objeciones contra esta corriente de trabajos desde el punto de vista científico y filosófico y concluir en la imposibilidad de aportar certeza objetiva sobre la emergencia de consciencia en máquinas, se expondrá una sección donde se analizará el motivo por el que el problema mente-cerebro se nos resiste tanto y por qué un punto de vista humanista orbitando al misterianismo sobre la consciencia fenoménica es, *a priori*, la corriente filosófica sobre la consciencia fenoménica que en opinión del autor tiene mayor evidencia de ser un enfoque razonable. Tras este análisis de la consciencia y la postura defendida sobre la misma se está en condiciones de aportar el argumento contra la descarga de la consciencia fenoménica en una máquina o, en otras palabras, aportar un argumento

que derribe el mito transhumanista de que se pueda poseer alguna garantía de tener una certeza de que, algún día y suponiendo que la ciencia se desarrolla en su plenitud, se pueda transferir la consciencia fenoménica de un individuo a un ordenador. Esta sección no solo aportará las premisas del argumento, sino que expondrá y rebatirá las objeciones que el materialismo reduccionista puede potencialmente esgrimir contra este.

2. Consciencia en máquinas. Funcionalismo

Pese a las limitaciones comentadas, el campo de la inteligencia artificial ha explorado la consciencia artificial (Chella & Manzotti, 2013). En concreto, la plausibilidad y el beneficio que se pueden obtener por implementar determinadas características de la consciencia en ordenadores mediante arquitecturas cognitivas (Garrido-Merchán *et al.*, 2020). Este intento tiene también su apoyo en la comunidad de filosofía de la mente, a través de la teoría del funcionalismo (Levin, 2004) y del conexionismo (Garson, 1997) que heredan del mecanicismo o materialismo reduccionista, sin ánimo de exhaustividad y ciñéndonos a lo relevante en el tema que nos aborda, la asunción, a juicio del autor de este artículo tremendamente antropocéntrica, e incluso no ausente de cierta soberbia, de que cualquier fenómeno o realidad presente en el universo puede ser explorada, entendida y replicada por el método científico y la tecnología desarrollada por el ser humano. En concreto, el funcionalismo es una teoría materialista que defiende que los estados mentales pueden ser representados mediante estados funcionales (Levin, 2004). Por tanto, la mente se puede replicar en un ordenador y la consciencia sería un epifenómeno que se replicaría en el propio ordenador (Chalmers, 2002). En concreto, la posición funcionalista se apoya en la idea de Hilary Putnam, que afirma que los estados mentales del cerebro pueden ser traducidos en estados funcionales en su totalidad. Y es de estos estados funcionales de donde la consciencia emerge (Putnam, 1992). En la misma línea de pensamiento, Rudolf Carnap, como positivista lógico, argumenta que cualquier actividad con sentido que se produce en la mente debe poder traducirse en un comportamiento externo, que en el caso de una inteligencia artificial puede ser o bien en movimientos de un robot o bien en el lenguaje en el que se expresa (Carnap, 1959). Esta creencia, ya que no se puede validar por el método científico carece en mi opinión de objetividad, ha incentivado a la comunidad de consciencia en máquinas a implementar modelos de computación con la expectativa de que de alguna manera vayan a ser conscientes simplemente por comportarse como un humano y simular este comportamiento mediante funciones. Aunque en mi opinión no se debe establecer una analogía entre inteligencia y consciencia, como se desarrolla en una sección posterior, donde se ilustran también las graves consecuencias de adoptar esta postura, incluso aceptando este postulado como cierto, la postura de Putnam fue indefendible desde el punto de vista técnico. En este sentido, la denotada GOFAI (Good Old Fashioned Artificial Intelligence [Jackson, 2019]) propuso sistemas expertos basados en lógica de primer orden que mediante

una base de conocimiento, que incluye una gramática de un lenguaje en particular y reglas para simular un programa de pregunta y respuesta y un motor de inferencia, pudiera mantener una conversación con un ser humano. Sin embargo, y pese a que este sistema teóricamente podría llegar a emular perfectamente una conversación con un ser humano, no se puede decir que por ello sea consciente. Sin embargo, un sistema de estas características en ningún caso sería capaz de emular el comportamiento de un ser humano, ya que no se puede adaptar a un conocimiento nuevo sin que previamente sea introducido como hecho en la base de conocimiento ni establecer analogías que se puedan obtener por el contexto. Por ejemplo, el ser humano tiene la capacidad de inferir que dos significantes similares se refieren al mismo concepto. Este fenómeno se da con mucha frecuencia en las conversaciones que se mantienen en redes sociales. En ellas, por ejemplo, se puede encontrar un gran número de formas de expresar el término «whisky», como «wiski» o incluso «guisqui». Dado un nuevo significante para referirse al significado de «whisky», los seres humanos podemos inferir fácilmente que nos referimos a este concepto por la fonología y por el contexto en el que se presenta el significante pero un sistema experto no contempla estos factores o tiene mucha dificultad para reconocerlos o adaptarse a nuevo conocimiento. Luego, si nos limitamos a juzgar si una entidad es consciente por su analogía en el comportamiento con los seres humanos, la inteligencia artificial clásica incluso refinada al extremo presentaría problemas detectables incluso por un test tan básico como el propuesto por Turing (Oppy & Dowe, 2003).

Más extrema es la teoría del conexionismo, que matiza al funcionalismo que estos estados mentales serían reproducidos por las conexiones de redes neuronales artificiales profundas complejas (Garson, 1997). Dada la alta expectativa presentada por esta corriente en las redes neuronales profundas artificiales, es adecuado definir técnicamente el concepto de red neuronal artificial antes de proceder al análisis de la plausibilidad de esta hipótesis. Una red neuronal artificial es un algoritmo de aprendizaje automático, como un árbol de decisión o una máquina de vector soporte. En general, los algoritmos de aprendizaje automático se componen de un conjunto, generalmente un tensor, de parámetros, \mathbf{W} , cuyos valores se ajustan según un conjunto de datos, $\mathbf{D} = \{\mathbf{X}, \mathbf{y}\}$ donde \mathbf{X} es una matriz cuyas columnas representan las características de cada instancia y cada fila representa a una instancia \mathbf{x}_i e \mathbf{y} es un vector que representa la variable que se va a predecir, cuyos elementos y_i son los valores respectivos de dicha variable para cada instancia \mathbf{x}_i . Por ejemplo, una instancia puede representar el peso y altura de cada individuo y la variable que se va a predecir puede ser si tiene sobrepeso o la probabilidad de desarrollar diabetes. A partir de un conjunto de datos etiquetado con la variable \mathbf{y} entrenamos al algoritmo, que ajusta los valores \mathbf{W} , para hacer predicciones de instancias no etiquetadas. Los parámetros \mathbf{W} se ajustan para optimizar el error cometido al realizar predicciones sobre datos de los que se desconoce la variable que se va a predecir (Murphy, 2012). En el caso de una red neuronal artificial, los parámetros representan los pesos de cada neurona de la red, y están relacionados entre sí mediante distintas arquitecturas especialmente pensadas para distintos tipos de problemas. En la figura 1 se ilustra la arquitectura de una red neuronal artificial totalmente conectada. Cada círculo representa una neurona y cada flecha

transmite la información de la neurona y tiene asociado un parámetro cuyo valor representa su peso. Una nueva instancia se inserta en la primera capa, columna de neuronas, y sus valores se van transformando según los pesos de la red y otros parámetros hasta la capa de salida, que suele representar la probabilidad de cada valor de la variable que se va a predecir, en el caso de ser una variable de valores categóricos, o el valor predicho de esa variable, en el caso de ser una variable de valores reales.

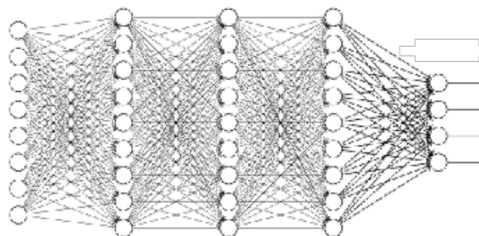


Figura 1. Representación de una red neuronal artificial profunda totalmente conectada.

Cada conjunto de valores de los parámetros \mathbf{W} representan hipótesis sobre los datos. La probabilidad de cada una de estas hipótesis se ajusta en función del procesamiento del conjunto de datos observado hasta que el algoritmo, generalmente, converge hacia una serie de valores que se considera óptima según los datos procesados y el método de entrenamiento empleado. Estos algoritmos de aprendizaje automático tienen a su vez una serie de hiperparámetros \mathbf{H} que flexibilizan el espacio de hipótesis que se van a contemplar sobre la variable a predecir con base en los datos observados. La peculiaridad de la red neuronal es que se trata de un aproximador universal de funciones si está bien hiperparametrizada (LeCun *et al.*, 2015). En concreto, un aproximador universal de funciones tiene la propiedad teórica de que puede modelizar cualquier función analítica suave hasta un nivel de precisión deseado. Podemos usar los algoritmos de aprendizaje automático para realizar predicciones basadas en datos vistos o se pueden reajustar permanentemente sus parámetros para que vayan adaptando sus predicciones en función de los datos que se introducen de forma continua, a lo cual se denomina aprendizaje por refuerzo. Las redes neuronales, mediante distintas formas de relacionar estos pesos, pueden generar predicciones de todo tipo de datos: imágenes, audio, vídeo, etc. Sin embargo, las redes neuronales profundas, y por extensión cualquier algoritmo de aprendizaje automático, no son más que modelos de funciones analíticas que se emplean para resolver problemas de regresión, clasificación o generación de datos. Estos algoritmos, por su carácter computacional, son emulables por una máquina de Turing universal pero no difieren en absoluto de cualquier otro algoritmo implementable en un ordenador en lo que respecta a sus propiedades ontológicas. Se puede argumentar en favor de la potencial consciencia de estos sistemas que estos algoritmos son capaces de adaptarse a nuevos escenarios mediante el aprendizaje por refuerzo, y potencialmente podrían mediante refinamiento técnico y la suficiente computación hacerse pasar por humanos superando el test de Turing e incluso resolver

problemas mediante métodos inaccesibles por la razón humana (como el caso de la red neuronal del sistema AlphaGo, descubriendo estrategias en el complejo juego Go desconocidas previamente [Holcomb *et al.*, 2018]). Sin embargo, estos algoritmos no poseen una intencionalidad propia ni libre albedrío, ya que su comportamiento está completamente determinado por el conjunto de valores que poseen sus parámetros e hiperparámetros (Murphy, 2012; LeCun *et al.*, 2015). Por otro lado, asociar consciencia fenoménica con inteligencia es problemático, como se detallará en la siguiente sección, ya que, entre otros motivos, la primera es irreductible, indefinible, subjetiva, cualitativa y relativa al ser humano, y la segunda puede ser, muy discutiblemente pero fuera del ámbito de estudio de este trabajo, objetivizada en cualquier sistema. Al no tener intencionalidad propia y ser la consciencia un fenómeno *a priori* independiente de la inteligencia, no hay ningún argumento objetivo para asociar la consciencia fenoménica a ningún sistema que implemente redes neuronales artificiales profundas.

La comunidad de consciencia artificial, sin embargo, toma las asunciones del funcionalismo y el conexionismo para proponer sistemas o robots que desarrollen qualia mediante la implementación de sistemas expertos, modelos de aprendizaje automático, metodologías híbridas u otras variantes de sistemas de procesamiento de la información que, en cualquier caso, son emulables mediante máquinas de Turing (Gámez, 2018). En concreto, la comunidad toma los modelos cognitivos publicados por neurocientíficos y los implementa en un ordenador para comprobar empíricamente si se obtiene algún beneficio en el rendimiento de los sistemas que incorporan conceptos pertenecientes a la consciencia o si se puede hipotetizar que son conscientes (Garrido-Merchan *et al.*, 2020). Por ejemplo, implementar el modelo neurocientífico de trabajo global en un programa informático para la toma de decisiones de un robot simulado por *software* (Garrido-Merchan *et al.*, 2020). Más aún, se han propuesto medidas para la consciencia que puede emerger en un ordenador, como la teoría de la información integrada (Tononi, 2015) o las pruebas de esquema de Winograd (Levesque *et al.*, 2012). Sin embargo, estas y otras medidas propuestas por esta comunidad son epistemológicamente objetivas y no pueden decir nada de un fenómeno ontológicamente subjetivo y cualitativo como es el caso de la consciencia fenoménica, debido a los argumentos planteados. Por tanto, estos tests miden la cantidad de información de un sistema o su capacidad de razonamiento en problemas generales y complejos, pero no pueden, sin tomar una gran cantidad de asunciones propias del funcionalismo, afirmar con certeza en ningún caso la aparición de consciencia fenoménica.

Del análisis presentado en esta sección se concluye que no se puede asegurar que la consciencia fenoménica emerja del procesamiento de datos por parte de un modelo de computación independientemente de su complejidad técnica, de la cantidad de información procesada, de su implementación en un *hardware* particular y de su capacidad para emular el comportamiento humano. Por ello, en la siguiente sección estudiaremos desde un punto de vista más amplio el fenómeno de la consciencia fenoménica y, en particular, el motivo por el cual es tan difícil extraer conclusiones sobre el problema mente-cerebro en lo que respecta a la consciencia fenoménica.

3. ¿Por qué el problema mente-cerebro se nos resiste tanto? Misterianismo

Como se ha expuesto, la consciencia es un fenómeno ontológicamente subjetivo y por tanto no puede ser medido directamente, ya que una medición es un enfoque epistemológicamente objetivo (Searle, 1997). Por tanto, no es directamente medible por el método científico (Timmermans y Cleeremans, 2015). En palabras de Aristóteles, la consciencia fenoménica se trata de un accidente de cualidad, y no de cantidad. La sustancia, nuestra forma ideal, por tanto, se trata de un fenómeno irreductible que no se podría expresar matemáticamente. Esto nos enfrenta al problema de cómo comprobar hipótesis y extraer conclusiones de nuestra experiencia consciente usando el método científico (Williams, 1934), que, por otro lado, tanto fruto nos ha dado en múltiples áreas del saber. Un enfoque efectivo es estudiar los fenómenos correlacionados con nuestra experiencia consciente (Koch *et al.*, 2016). Se pueden discriminar mediante pruebas fMRI qué áreas del cerebro están correlacionadas con una experiencia consciente (Harrison *et al.*, 2013) o identificar comportamientos que un individuo efectúa correlacionados con su experiencia consciente (Garrido-Merchan y Molina, 2020). Incluso se puede predecir si un sujeto ha sido consciente o no de una imagen expuesta en un intervalo de nanosegundos enmascarándola con otra imagen antes y después. No obstante, dichas mediciones se aplican a fenómenos correlacionados con la consciencia pero no determinan cómo y por qué emerge la propia consciencia en el ser humano (Chalmers, 2004). Por otro lado, como es bien sabido, correlación no implica causalidad (Pearl, 2009). Por tanto, la actividad neuronal puede ser una correlación espuria del potencial fenómeno que haga emerger la consciencia fenoménica. Por ejemplo, se posee evidencia y testimonios de sujetos que han experimentado consciencia en estados de coma (Teasdale y Jennett, 1974) o que son conscientes con la mayor parte de su cerebro dañado (Feuillet *et al.*, 2007). Con tal evidencia, cobra peso la hipótesis de que la correlación entre la emergencia de consciencia y la actividad cerebral detectable mediante fMRI pueda originarse debido a una causa oculta que explique ambos hechos correlacionados (Murphy, 2012). Un ejemplo de correlación espuria se encuentra en el aparentemente muy poco probable postulado «comer helado causa asesinatos» (Peters, 2013). En este estudio, se demostró que la causa de los asesinatos no era comer helado sino un tiempo críticamente caluroso. Las altas temperaturas no solo influían en los actos de los asesinos sino que, al mismo tiempo, provocaban que la gente consumiese helado. Pese a haber una alta relación entre los crímenes y el consumo de helado, no existe ninguna relación causal, sino que ambos fenómenos comparten una causa común, las altas temperaturas, y son independientes entre sí. Del mismo modo, no podemos tener certeza objetiva de que la actividad cerebral sea la causa de la consciencia, sino que, por todos los argumentos presentados, parece haber una causa no estudiable por el método científico que provoca ambos efectos.

Dada esta imposibilidad de poseer un método para medir la consciencia directamente y la alta incertidumbre que cualquier postulado científico conlleva a la hora de explicar la consciencia

cia fenoménica, una postura que ha adquirido popularidad es el eliminativismo o ilusionismo (Frankish, 2016), popularizado por Daniel Dennett (Dennett, 2016). Esta teoría de la mente equipara la consciencia de uno mismo a una ilusión, sugiere que la consciencia es una ilusión creada por la mente humana y que no existe a nivel ontológico en el universo (Frankish, 2016). En otras palabras, niega su existencia o, en su versión más moderada, la reduce a un epifenómeno (Chalmers, 2002). Sin embargo, como ya se ha argumentado, no podemos refutar la existencia de la consciencia ya que un ejercicio de introspección nos brinda la certeza de estar vivos, de existir. En palabras de Descartes en su segunda meditación (Descartes, 2013), «*cogito ergo sum*», la que en opinión del autor es la realidad más evidente y absenta de discusión de la que un observador particular puede dar crédito. Negar la existencia de la consciencia al no ser modelizable por el método científico es una postura propia del reduccionismo materialista, que asume el determinismo como propiedad del universo que nos rodea y eliminar cualquier fenómeno que no pueda ser trabajado por el método científico genera los problemas en ciencia natural descritos por Thomas Nagel en *La mente y el cosmos* (Nagel, 2012).

En mi opinión, defender una postura ilusionista es extremadamente complicado. Ser ilusionista implica negar nuestra propia esencia (Dennett, 2016), en términos aristotélicos. Implica introducirnos en la caverna de Platón al solo reconocer como realidades ontológicas del universo las perceptibles por el ser humano, lo cual sería equivalente a solo ser testigos de las sombras y no del mundo de las ideas. Implica negar la pregunta por el sentido, abandonar cualquier postura teleológica, es decir, negar el propósito o fin de cualquier objeto y ser. Implica por tanto negar al teísmo, ya que la quinta vía que santo Tomás de Aquino propone para la existencia de Dios pasa por un argumento teleológico. Implica, por tanto, negar en consecuencia la existencia ontológica de causas finales propuesta por Aristóteles. Si negamos la existencia de la consciencia, nos encontramos con el problema de que deberíamos negar nuestra experiencia de la realidad que nos rodea, y precisamente si de algo no tenemos ninguna duda es de que experimentamos algo. Aunque ese algo sea un fenómeno cualitativo y ontológicamente subjetivo (Searle, 1997) y aunque todo lo que percibiésemos fuera una completa ilusión creada por el demonio de Descartes (Gillespie, 2006) o similar a la realidad plasmada en la película *Matrix* (Wachowski, 1999). Pese a todo ello, somos conscientes de nuestra identidad, somos conscientes de una serie de sensaciones en cada momento y experimentamos estas sensaciones (Blackmore, 2014). A los problemas asociados a negar todos los postulados metafísicos mencionados anteriormente se le suma que adoptar una postura ilusionista implica asumir un pensamiento antropocéntrico al aceptar, no sin cierta soberbia, que el ser humano, ayudado por el método científico y la tecnología, es capaz de discriminar lo que es real y lo que no. Si un fenómeno, como la consciencia fenoménica, no es trabajable por el método científico, entonces es descartado. Si en vez de tomar tal asunción, el filósofo hace un ejercicio de humildad, puede reconocer sin complejos la existencia ontológica de realidades en el universo de las cuales el ser humano no puede obtener conocimiento objetivo epistémico. Dichas realidades ontológicas son, por tanto, modelizables por la metafísica pero misterios desde un punto de vista objetivo (McGinn, 1989). En particular, una de las modelizaciones que se puede adoptar para la consciencia fenoménica es una postura dualista.

Sin embargo, el dualismo intervencionista de Descartes (Descartes, 2013) presenta el problema de cómo puede una realidad no física intervenir en una realidad física y viceversa. Para solucionar esta cuestión, se propuso que la consciencia fenoménica podía ser una suerte de epifenómeno, un producto accidental del cuerpo humano que surge de su actividad pero no interviene en su comportamiento (Chalmers, 2002). Sin embargo, el dualismo epifenoménico presenta una objeción difícil de refutar en el argumento expuesto en la breve fábula titulada *Un dualista desgraciado*, de Raymond Smullyan (Smullyan, 1981). En este escrito, se plasma el problema de que si el alma no interviene en el cuerpo, entonces, si esta es descartada, el comportamiento expuesto del cuerpo debería verse inalterado. Sin embargo, si el comportamiento es inalterado el sujeto seguirá hablando de conceptos como alma, esencia o consciencia fenoménica, lo cual no tiene sentido ya que en teoría, al no tener alma, debería ser un *zombie* filosófico que no pudiera inferir este tipo de nociones.

Según los argumentos anteriormente expuestos y sumado a que el método científico no puede trabajar con esta realidad, nos encontramos con gran evidencia de que el concepto de consciencia fenoménica puede ser una realidad ontológica que escapa de nuestra comprensión. No solo de la comprensión que podemos llevar a cabo como seres humanos sin la ayuda de tecnología, al más puro estilo de la mente humeana, sino que tampoco podemos acceder a este conocimiento mediante el uso del método científico y la tecnología. Al no encontrar consenso filosófico en torno a la cuestión (Chalmers, 2002) y no poder adquirir conocimiento objetivo epistémico de esta realidad que, por otro lado, es evidente, ello implica que es razonable, dada la alta evidencia de esta hipótesis o, mejor dicho, la baja evidencia del resto de hipótesis, adoptar una postura misterianista (Kriegel, 2003), según la cual la consciencia es un concepto que supera nuestras actuales capacidades cognitivas, es decir, un concepto que nos está cerrado cognitivamente, y un misterio que nos resulta inexplicable (Arana, 2015). Según esta postura, independientemente de los avances que la ciencia puede otorgar, la incertidumbre epistémica, dada nuestra ignorancia sobre las causas, mecanismos o leyes que producen este fenómeno y que no son estudiables por el método científico, que tenemos acerca de la naturaleza de la consciencia fenoménica siempre será máxima y nunca podrá reducirse.

Asumiendo una postura misterianista, dados los argumentos propuestos, sobre el fenómeno de la consciencia fenoménica, nos encontramos con una implicación directa. Dado que la consciencia es un fenómeno ontológico subjetivo, cualitativo y relativo al observador, desde mi punto de vista, no puedo afirmar de forma tajante y objetiva que cualquier otro ser sea consciente. Bertrand Russell escribe en *Analogía* (Russell, 1970) que lo único que podemos llegar a afirmar, de forma subjetiva, es que si otro ser es análogo a mí, entonces es más probable que sea consciente. En otras palabras, trabajando con creencias subjetivas al más puro estilo bayesiano, podemos introducir distribuciones *a priori* acerca de la probabilidad de que un ser sea consciente o no dadas las características que percibimos de dicho ser por nuestros sentidos o mediante la tecnología. Sin embargo, adoptar esta postura que *a priori* puede resultar muy moderada, en mi opinión, presenta problemas muy importantes y consecuencias muy serias. ¿A qué nos referimos con que un ser es análogo a mí? Si nos referimos a la semejanza

física externa, y considerando el concepto de consciencia fenoménica, ¿es más probable que una persona físicamente sana sea consciente a que lo sea una que ha perdido sus extremidades?, ¿es más probable que una persona adulta sea consciente con respecto a un embrión? Si nos referimos a la semejanza en términos de comportamiento, ¿es más probable que una persona neurotípica sea consciente con respecto a una persona con síndrome de Down o trastorno del espectro autista? En mi opinión, no nos corresponde emitir tal juicio y deberíamos olvidar esta idea ya que, como el lector ya habrá pensando, esta asunción conlleva consecuencias éticas y políticas catastróficas. Por tanto, un sujeto consciente no puede decir nada objetivo sobre la emergencia de consciencia fenoménica en otro ser y, en consecuencia, no podría ni siquiera afirmar objetivamente que si se replica a la perfección su cerebro en un ordenador su consciencia emergerá en dicho artefacto. En la siguiente sección argumentamos de forma más detallada esta idea explorando las consecuencias que este hecho tiene en el transhumanismo y argumentando las objeciones que distintas posturas filosóficas pueden tener con respecto a la imposibilidad de poder afirmar objetivamente mediante el método científico que se puede descargar la consciencia en un ordenador.

4. Argumento contra la descarga de la consciencia fenoménica en una máquina. Objeciones

El transhumanismo es un movimiento cultural e intelectual internacional que tiene como objetivo final transformar la condición humana mediante el desarrollo y fabricación de tecnologías ampliamente disponibles, que mejoren las capacidades humanas, tanto a nivel físico como psicológico o intelectual. Una de las metas futuras de este movimiento radica en transferir la consciencia fenoménica de un individuo a un ordenador. Sin embargo, y tal como se ha argumentado en las secciones previas, dicha pretensión se asimila más a un mito que a una meta científica, ya que la consciencia fenoménica no es objeto de estudio del método científico. Sin embargo, en la actualidad no es raro observar eslóganes publicitarios anunciando que una inteligencia artificial siente, o encontrar en el cine y la literatura constantes referencias a la transferencia de la consciencia del ser humano a una máquina. Tales fenómenos se están introduciendo en la psicología popular, y, por tanto, es crítico enfatizar que su cercanía al mito es infinitamente superior a su viabilidad práctica. Si el ser humano, desde un punto de vista transhumanista, anhela persistir su consciencia en una máquina, es importante advertir de la imposibilidad de asegurar que este fenómeno es viable desde un punto de vista objetivo. Ya que incluso en el remoto caso de que creásemos un ser cognitivamente semejante a nosotros, pero en el cual no podemos introducir nuestra consciencia, no será nada más que lo llamado como *zombie* filosófico, una inteligencia artificial hecha a nuestra imagen y semejanza pero no consciente (Chalmers, 2004). Supóngase

que una persona accede, mediante un supuesta tecnología futura, a descargar su consciencia en una máquina, aunque esta máquina replique el comportamiento de esa persona a la perfección, jamás podremos saber si esa entidad es consciente o no desde un punto de vista objetivo. Y mucho menos asegurar que la consciencia fenoménica de ese artefacto corresponde a la de la persona que quiso descargar su consciencia en ese artefacto. Por tanto, el riesgo al que se somete la persona que, persuadida de que una tecnología es capaz de transferir su consciencia a una máquina, accede a transferir su consciencia a una máquina es la total incertidumbre de que efectivamente el proceso sea satisfactorio. Dado que la consciencia fenoménica es una propiedad fuera de nuestro alcance, jamás podremos saber objetivamente si la transferencia ha sido satisfactoria sin asumir una postura funcionalista. Por tanto, el estudio realizado de la consciencia fenoménica y la postura misterianista adoptada en torno a este fenómeno tiene como propósito proponer el siguiente argumento contra la afirmación de que la descarga de la consciencia fenoménica en una máquina se puede efectuar con garantías.

Para poder descargar la consciencia fenoménica de un individuo en una máquina, como por ejemplo en un ordenador, necesitamos medir de forma objetiva la consciencia fenoménica que pudiera emerger de dicha máquina u ordenador. Por otro lado, y como ya se ha argumentado en secciones previas, la consciencia fenoménica es un fenómeno subjetivo, cualitativo, privado y relativo a un observador. En cuanto al método científico, aporta certeza sobre un hecho mediante medidas objetivas verificadas empíricamente, es decir, fenómenos que se pueden medir objetivamente. Si aceptamos estas premisas, de aquí se sigue que, para poder verificar si una máquina es consciente, se necesita verificarlo empíricamente, es decir, medir objetivamente la consciencia fenoménica que emerge de la máquina u ordenador. Tal y como se ha expuesto, no se puede verificar que una máquina es consciente con el método científico, al menos tal cual lo conocemos. De aquí se infiere lógicamente que la ciencia no aporta una certeza de que fuese viable la descarga de nuestra consciencia fenoménica en un ordenador, independientemente de su desarrollo futuro. La única certeza que se puede tener es la total incertidumbre, es decir, la máquina podría ser consciente o no y nunca podremos saberlo.

El argumento expuesto afirma tajantemente la imposibilidad de obtener una certeza sobre la transferencia de consciencia fenoménica de un individuo a una máquina. Ante tal afirmación tajante, diferentes corrientes filosóficas pueden presentar las siguientes objeciones, que se responden desde el punto de vista misterianista sobre la consciencia fenoménica.

En primer lugar se puede objetar que se han obtenido correlaciones de la actividad cerebral mediante fMRI, con reportes en los que el sujeto se declara consciente. Luego la consciencia puede ser trabajada por el método científico. Aun en este caso, no sabemos si el sujeto es un *zombie* filosófico. La consciencia no es un fenómeno objetivo, es subjetivo, relativo y propio de cada individuo. Correlación no es causalidad, como se ha expuesto. La causa de la actividad cerebral y de la consciencia puede estar oculta al método científico y explicar ambas, no podemos saberlo. Las fMRI miden un fenómeno físico pero la consciencia fenoménica, y no de acceso o la del yo, es totalmente subjetiva, y por tanto no es medida directamente por las fMRI. Luego por una correlación no se puede concluir una relación causal.

No obstante, si se detecta una correlación perfecta entre una onda del cerebro con fMRI y la observación externa de la consciencia de un sujeto, ¿no se podría afirmar que el cerebro genera esa consciencia y que la hemos medido? De nuevo, se habla de correlación, no de causalidad. Puede haber otra causa que explique la onda y la consciencia; por ejemplo: una dimensión sobre la cual no podemos interferir que genera esa onda en el cerebro y donde se alberga la consciencia fenoménica.

Una objeción común es apelar al avance y los hitos de la ciencia para inferir por inducción que cualquier fenómeno será explicado por la ciencia. Sin embargo, este es un argumento claramente cientificista. El propio método científico no puede demostrar que el método científico es el único método válido para explicar todos los fenómenos que percibimos. Es muy antropológico pensar que todo el universo puede ser explicado por un humano. La epistemología puede no explicar el conjunto total de la ontología, que puede reservar conceptos cerrados al conocimiento humano como es el de la consciencia fenoménica.

El ilusionismo propone a esta imposibilidad de estudiar la consciencia fenoménica que esta no existe, es producto de una definición errónea, una ilusión (Dennett, 2016). No existen realidades subjetivas sino solo objetivas. Un simple ejercicio de introspección basta para sentir que la consciencia fenoménica es, con seguridad, el único fenómeno del que una persona puede estar completamente seguro de su existencia. Es imposible saber si la percepción de la realidad procesada por el cerebro es ficticia. Pero la percepción de dichos estímulos internos y externos por la consciencia fenoménica es una realidad de la que cada individuo puede dar testimonio.

Ante este problema, una postura más moderada afirma que la consciencia es un epifenómeno. Luego no hace falta usar la ciencia para afirmar su existencia. Por analogía con nosotros, emergerá en un sistema igual formado. Esta hipótesis presenta evidencias en contra que hacen que su probabilidad de ser cierta sea reducida: si la consciencia es un epifenómeno, entonces no puede interactuar con esta realidad. Luego se niega la intencionalidad de las acciones y el libre albedrío, lo cual resulta tremendamente problemático a nivel objetivo y subjetivo. Ya se analizó en la sección anterior que recurrir a la objeción del behaviorismo o funcionalismo por la cual si se comporta como X entonces es X o recurrir a analogías presenta los problemas que se han objetado contra la *Analogía* de Russell. Una última objeción consiste en decir que si nosotros somos conscientes y consideramos a nuestro cuerpo como una máquina, entonces una máquina también podrá ser consciente. Ante esta afirmación, simplemente se dirá que nunca podremos saber si una máquina es consciente o no por los argumentos expuestos y que si tenemos una certeza es que el ser humano trasciende por su dimensión subjetiva, por su intencionalidad, por su libre albedrío y por su pregunta por el sentido a cualquier máquina que podamos conocer.

Conclusiones

Pese a la gran capacidad de resolución de problemas y la aparente semejanza que una inteligencia artificial puede tener hoy en día con un ser humano, no podemos argumentar con

certeza que esta sea consciente. Por extensión, tampoco podemos afirmar que nosotros podamos descargar nuestra consciencia en un ordenador. En lo respectivo al transhumanismo esto supone un problema, puesto que afirmar que no se puede argumentar con certeza que podamos persistir en un ordenador implica que seguiremos indefinidamente siendo dependientes de nuestro cerebro biológico para ser conscientes de nosotros mismos. O, como mínimo, de una secuencia de ADN transformada por ingeniería genética que haga nuestro cerebro biológico más avanzado, pero, al fin y al cabo, biológico o derivado del ADN que nos ha creado. En cualquier caso, no creado desde cero en un artefacto similar a un ordenador. En conclusión, si pensamos en el transhumanismo parece una alternativa más segura explorar la ingeniería genética o los implantes artificiales que la idea de persistencia o trascendencia en un sistema físico independiente de nuestro cerebro biológico. Como mucho, podríamos explorar la idea de embeber a nuestro cerebro en sistemas artificiales. La corriente transhumanista debe abandonar y publicitar la idea, o más bien el mito, de la viabilidad objetiva de la transferencia de la consciencia fenoménica en un ordenador y, en un ejercicio de humildad antropológica, conceder que el universo que nos rodea presenta fenómenos que están fuera del alcance del método científico, tal y como hoy lo conocemos.

Agradecimientos

Se agradece la financiación del Plan Nacional Español I+D+i, beca TIN2016-76406-P. También se agradece la financiación de PID2019-106827GB-I00 / AEI / 10.13039/501100011033.

Bibliografía

- Descartes, R. (2013). *Meditations on first philosophy*. Broadview Press.
- Bering, J. M. (2006). The folk psychology of souls. *Behavioral and brain sciences*, 29(5), 453.
- Block, N. (1995). Some concepts of consciousness. *Sciences*, 18(2).
- Feuillet, L., Dufour, H. y Pelletier, J. (2007). Brain of a white-collar worker. *The Lancet*, 370(9583), 262.
- Searle, J. (1997). *The Mystery of Consciousness*. *The New York Review of Books*, Nueva York.
- Timmermans, B. y Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. *Behavioural methods in consciousness research*, 21.
- Williams, D. C. (1934). Scientific method and the existence of consciousness. *Psychological Review*, 41(5), 461.
- Koch, C., Massimini, M., Boly, M. y Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, 17(5), 307-321.
- Harrison, A. H. y Connolly, J. F. (2013). Finding a way in: a review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. *Neuroscience & Biobehavioral Reviews*, 37(8), 1403-1419.
- Merchán, E. C. G. y Molina, M. (2020). A machine consciousness architecture based on deep learning and gaussian processes. *In International Conference on Hybrid Artificial Intelligence Systems* (pp. 350-361). Springer, Cham.

- Chalmers, D. J. (2004). How can we construct a science of consciousness?
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, 304(7872), 81-84.
- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12), 11-39.
- Chalmers, D. J. (2002). *Philosophy of mind: Classical and contemporary readings*.
- Gillespie, A. (2006). Descartes' demon: A dialogical analysis of Meditations on First Philosophy. *Theory & psychology*, 16(6), 761-781.
- Wachowski, A., Wachowski, L., Reeves, K., Fishburne, L., Moss, C. A., Weaving, H., ... y Staenberg, Z. (1999). *Matrix*. Burbank: Warner Home Video.
- Blackmore, S. (2014). *Zen and the Art of Consciousness*. Simon and Schuster.
- Chalmers, D. J. (2004). *Zombies on the web*.
- Copleston, F. (2003). *History of Philosophy Volume 1: Greece and Rome (Vol. 1)*. A&C Black.
- Nelson, K. (2003). Narrative and the emergence of a consciousness of self. *Narrative and consciousness*, 17-36.
- Nour, M. M., Evans, L., Nutt, D. y Carhart-Harris, R. L. (2016). Ego-dissolution and psychedelics: validation of the ego-dissolution inventory (EDI). *Frontiers in human neuroscience*, 10, 269.
- Timmermann, C., Roseman, L., Williams, L., Erritzoe, D., Martial, C., Cassol, H., ... y Carhart-Harris, R. (2018). DMT models the near-death experience. *Frontiers in psychology*, 9, 1424.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
- Jackson, F. (2004). *There's something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument*. MIT Press.
- Nagel, T. (1974). What is it like to be a bat? *The philosophical review*, 83(4), 435-450.
- Solomon, P., Leiderman, P. H., Mendelson, J. y Wexler, D. (1957). Sensory deprivation: A review. *American Journal of Psychiatry*, 114(4), 357-363.
- Chalmers, D. (2007). The hard problem of consciousness. *The Blackwell companion to consciousness*, 225-235.
- Krauss, L. M. (2013). *A universe from nothing: Why there is something rather than nothing*. Atria Books.
- Chella, A. y Manzotti, R. (2013). *Artificial consciousness*. Andrews UK Limited.
- Garrido-Merchán, E. C., Molina, M. y Mendoza, F. M. (2020). An artificial consciousness model and its relations with philosophy of mind. arXiv preprint arXiv:2011.14475.
- Levin, J. (2004). *Functionalism*.
- Garson, J. (1997). *Connectionism*.
- Tononi, G. (2015). Integrated information theory. *Scholarpedia*, 10(1), 4164.
- Levesque, H., Davis, E. y Morgenstern, L. (2012, May). The winograd schema challenge. *In Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*.
- Vanney, C. E. y Juan, F. (2018). ¿Determinismo o indeterminismo? Universidad Francisco de Vitoria.
- Nagel, T. (2012). *Mind and cosmos: why the materialist neo-Darwinian conception of nature is almost certainly false*. Oxford University Press.
- Smullyan, R. M. (1981). *An unfortunate dualist*. HOFSTADTER, DR & DENNETT, DC *The Mind's I*. New York: Basic Books, 383-384.
- Dennett, D. C. (2016). Illusionism as the obvious default theory of consciousness. *Journal of Consciousness Studies*, 23(11-12), 65-72.
- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12), 11-39.
- Platón. *La República*. Libro VII.

- McGinn, C. (1989). Can we solve the Mind-Body problem? *Mind*, 98(391), 349-366.
- SANTO TOMÁS, D. A. (2001). Suma teológica. Edición dirigida por los Regentes de Estudios de las Provincias Dominicanas en España. Madrid: BAC, 5, 725-731.
- Aristóteles. Metafísica. Libro I.
- Arana, J. (2015). La conciencia inexplicada. Madrid: Biblioteca Nueva.
- Kriegel, U. (2003). The new mysterianism and the thesis of cognitive closure. *Acta Analytica*, 18(1), 177-191.
- Russell, B. (1970). *Analogy. Essays on other minds*, 3-8.
- Putnam, H. (1992). The nature of mental states. *The philosophy of mind: Classical problems/contemporary issues*, 51-58.
- Oppy, G., & Dowe, D. (2003). The turing test.
- Carnap, R. (1959). Psychology in physical language (pp. 170-173). na.
- Jackson, P. C. (2019). Introduction to artificial intelligence. *Courier Dover Publications*.
- Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT Press.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Holcomb, S. D., Porter, W. K., Ault, S. V., Mao, G., & Wang, J. (2018, March). Overview on deepmind and its alphago zero ai. In *Proceedings of the 2018 international conference on big data and education* (pp. 67-71).
- Gamez, D. (2018). Human and machine consciousness (p. 234). Open Book Publishers.
- [56] Peters, J. (2013). When ice cream sales rise, so do homicides. coincidence, or will your next cone murder you. Slate. Retrieved from slate.com/news-and-politics/2013/07/warm-weather-homicide-rates-when-ice-cream-sales-rise-homicides-rise-coincidence.html.

